

P11-2008-81

А. В. Нечаевский, В. В. Кореньков

МОДЕЛИРОВАНИЕ DATAGRID
С ПРИМЕНЕНИЕМ ПАКЕТА OPTORSIM

Нечаевский А. В., Кореньков В. В.

P11-2008-81

Моделирование DataGrid с применением пакета OptorSim

Рассмотрены основные задачи, решаемые системами моделирования DataGrid. Описываются пакеты OptorSim и GridSim. На примере моделирования сегмента DataGrid показано применение OptorSim для моделирования DataGrid и представлены полученные результаты.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2008

Nechayevskiy A. V., Korenkov V. V.

P11-2008-81

DataGrid Simulation with OptorSim Package

The work considers the main problems that are solved by DataGrid simulation systems. OptorSim and GridSim packages are described. The application of the OptorSim software package for DataGrid simulation is demonstrated on an example of modelling a DataGrid segment. The obtained results are presented.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2008

ВВЕДЕНИЕ

Самым масштабным научным проектом последних лет является создание в ЦЕРН большого адронного коллайдера (LHC — Large Hadron Collider) [1]. После запуска в нормальном режиме ускоритель будет производить около 15 петабайт данных ежегодно (15 миллионов гигабайт). Открытие новых фундаментальных частиц и выявление их свойств на ускорителе LHC невозможно без статистического анализа больших объемов данных, собираемых с детекторов LHC, и детального сравнения с данными теоретического моделирования. Задача проекта LCG (LHC Computing Grid) [2] — создать и поддерживать инфраструктуру для хранения и анализа данных, которой будет пользоваться все сообщество ученых, участвующих в LHC. Для решения этой грандиозной задачи разрабатывается новейшая распределенная модель хранения и обработки данных, основное место в которой занимают grid-технологии.

Суть распределенной модели состоит в том, что весь объем информации с детекторов LHC после обработки в реальном времени и первичной реконструкции должен направляться для дальнейшей обработки и анализа в региональные центры.

Понятно, что создание такой распределенной системы является крайне сложной задачей. Прежде всего, это необходимость обеспечения достаточной пропускной способности, поддержка различного оборудования, проблемы сохранности данных (устойчивость к повреждениям и удалениям) на протяжении всего жизненного цикла проекта LHC, обеспечение распределения ресурсов между различными группами пользователей. Для решения этих задач на этапе проектирования могут использоваться пробные системы. Возможности тестовых grid-систем, как правило, ограничены, а создание испытательного grid с размером, близким к реальным, является дорогим и трудоемким. Также невозможно проверить множество различных сценариев использования DataGrid на реальных испытательных стендах. Под сценарием использования DataGrid понимаются набор правил для пользователей и алгоритмы работы ресурсов. Поэтому возникает задача моделирования grid-систем как средства изучения сложных сценариев [3].

В процессе использования уже существующих grid возникают вопросы, связанные с внесением изменений в grid и тем, как эти изменения отразятся на работоспособности grid и ее отдельных сегментов. За длительное время мониторинга передач данных из ЦЕРН на российские сайты JINR (ОИЯИ),

SINP (МГУ), ИИЕР (ИФВЭ), ИТЕР (ИТЭФ) и другие сайты, включенные в структуру RDIG [4], выяснилось, что одна из основных проблем участка CERN–RDIG связана с ошибками, возникающими из-за ограничения времени передачи данных. Такие ограничения необходимы для контроля состояния элементов системы и организации механизма ограничения очередей. Возникает вопрос, как изменится ситуация при увеличении размеров передаваемых файлов и что надо сделать для увеличения пропускной способности канала.

В настоящее время существует несколько пакетов моделирования grid-систем. Наиболее известные из них — MicroGrid [5], SimGrid [6], OptorSim [7] и GridSim [8]. В большинстве своем эти программы моделируют вычислительные системы Grid, а возможность моделировать DataGrid-системы реализована лишь в некоторых из них, например, в OptorSim и GridSim. Результатом работы такой программы могут являться собранные в ходе моделирования статистические данные о наиболее важных характеристиках моделируемой системы: использование дисковых ресурсов, пропускной способности каналов, вероятности потерь данных и т. п.

В связи с ожидаемым запуском LHC в ЦЕРН и последующей пересылкой больших объемов данных такое исследование может оказаться полезным и послужить для аргументированных рекомендаций по улучшению оборудования, сети или используемых стратегий репликации данных.

ОСНОВНЫЕ ЗАДАЧИ, РЕШАЕМЫЕ СИСТЕМАМИ МОДЕЛИРОВАНИЯ DATAGRID

Система моделирования DataGrid может быть использована для решения трех основных типов задач:

- выявление «слабых мест» системы (перегруженные каналы, места возникновения больших очередей и прочие «узкие места»);
- возможность тестирования различных сценариев использования DataGrid;
- поиск оптимальной конфигурации ресурсов.

ТРЕБОВАНИЯ К СИСТЕМЕ МОДЕЛИРОВАНИЯ

1. Необходимо моделирование основных элементов DataGrid (ресурсов хранения данных (SE), брокеров ресурсов (RB), каталога реплик (RC), сети, пользователей (User) и сайтов);

2. Скорость работы модели должна значительно превосходить скорость работы реальной DataGrid;
3. Необходима статистика по отдельным элементам (например, использование дисковых ресурсов) и по работе модели в целом (время выполнения, количество переданных файлов, загрузка сети и др.);
4. Необходимо моделирование сбоев оборудования;
5. Результаты моделирования должны быть сопоставимы с реальной ситуацией.

ПАКЕТЫ МОДЕЛИРОВАНИЯ DATAGRID

OptorSim — пакет моделирования Grid, реализованный на языке Java. Он позволяет оценивать различные алгоритмы оптимизации и стратегии копирования. В системе grid каждый сайт может содержать несколько элементов хранения и/или вычислительных элементов. В пакете содержатся несколько конфигурационных файлов, которые используются для задания параметров модели. Подробное описание файлов и характеристик представлено в руководстве пользователя [9]. Создание конфигурационных файлов и доработка модели не вызывают особых трудностей. Существует возможность моделирования фоновых трафика. Одним из достоинств пакета является наличие графического интерфейса для визуального отображения модели. Статистика в OptorSim представлена как по каждому элементу в отдельности, так и по работе модели. Отчетная информация доступна в виде таблиц, графиков и диаграмм. Открытый доступ к исходному коду OptorSim дает возможность доработки системы самостоятельно.

GridSim позволяет моделировать различные классы гетерогенных ресурсов, пользователей, приложений, брокеров ресурса и планировщиков. Моделирование DataGrid — это одна из возможных областей применения пакета GridSim, предоставляющего возможность определять ресурсы с гетерогенными компонентами хранения. Система обладает гибкостью для осуществления различных стратегий управления данными [3]. На сайте разработчиков много примеров по работе с программой [8]. На начальном этапе создания модели может использоваться VisualModeler (автоматическое создание пользователей и ресурсов, генерация кода и т. п.). Однако завершающий этап — создание связей — необходимо выполнять вручную, что довольно трудоемко и может стать причиной возникновения ошибок. Стоит отметить возможность моделирования фоновых трафика и моделирование отказов ресурсов [10]. Набор инструментов GridSim успешно использовался для моделирования брокера ресурсов Grid-Nimrod-G [11]. GridSim позволяет собрать статистику по всем или отобранным операциям, что зависит от того, что конкретно интересует пользователя. Это может быть и мониторинг дискового пространства,

и размеры очередей, и загруженность сети, и многое другое. В самом простом случае можно выводить всю интересующую информацию в файл и после моделирования использовать утилиты работы с текстом (например, `grep` или `awk`), чтобы отобрать данные, необходимые для дальнейшего анализа. Для графического представления результатов можно использовать `Gnuplot`.

Применение этих систем для моделирования различных стратегий и исследования различных реализаций `grid` может существенно облегчить и ускорить процесс принятия решений. `OptorSim` проще в использовании и позволяет исследовать алгоритмы оптимизации и стратегии копирования, в то время как `GridSim` позволяет более детально и тщательно описать исследуемую модель, что в свою очередь требует дополнительного времени и навыков работы с системой. Таким образом, выбор конкретного средства моделирования зависит от поставленных задач и сроков их исполнения.

Для моделирования участка CERN–RDIG использовался пакет `OptorSim`.

ОПИСАНИЕ МОДЕЛИ

Участок CERN–RDIG является частью глобальной структуры LCG. На данный момент трафик между ЦЕРН и сайтами RDIG проходит через сеть GEANT2 [12]. Отметим, что через маршрутизаторы (роутеры) проходит трафик не только российских сайтов, но также и других участников проекта, которые для простоты не учитывались. Топология моделируемого участка сети представлена на рис. 1.

Моделировалась работа четырех сайтов RDIG: JINR (ОИЯИ), SINP (МГУ), ИНЕР (ИФВЭ), ИТЕР (ИТЭФ). В таблице ниже приведены их характеристики: число CPU и объем ресурсов хранения.

	JINR	ИНЕР	ИТЕР	SINP
Количество CPU*	330	80	84	48
Объем дисковых ресурсов, Тбайт**	64	1,12	25,1	15,5

Моделировался следующий сценарий работы `grid`. Все файлы изначально находились на ресурсах хранения ЦЕРН. Сайты RDIG инициировали запуск задач. Когда задача запущена, для каждого файла в задаче определяется наилучшее расположение и моделируется передача данных на сайт. Файлы копируются по одному, но несколько сайтов могут копировать файлы по

*Максимальное количество CPUs на сайтах в период с 04.2007 по 04.2008, <http://www.egee-rdig.ru/>

**Максимальное доступное место на дисках в период с 04.2007 по 04.2008, <http://www.egee-rdig.ru/>

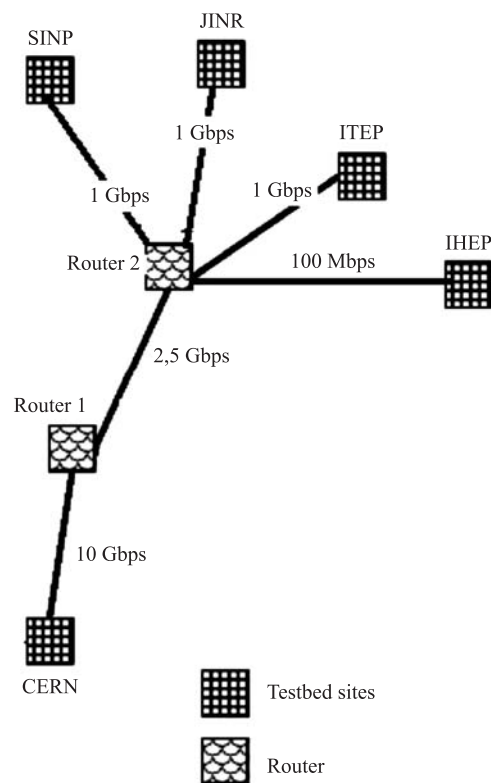


Рис. 1.

одной сети одновременно. Неудобство в том, что для каждого файла необходимо ждать, пока закончится передача предыдущего файла, в то время как реально сервис передачи данных (FTS) [14, 15] может выполнять копирование нескольких файлов одновременно. Однако описанная ситуация тоже вполне вероятна.

OptorSim позволяет моделировать фоновый трафик. Эта возможность использовалась в модели для создания дополнительных нагрузок на каналах передачи данных.

Пользователи запускали задачи по распределению Гаусса с пиком в 15 ч. Предполагалось, что исходные файлы, размером от 1 до 5 Гбайт находились на сайте ЦЕРН. Каждая задача могла содержать по несколько файлов (от 5 до 10). Эти характеристики близки к реальным, где размер передаваемых файлов приблизительно 2,6 Гбайт и может несколько варьироваться в зависимости от задачи.

Пропускная способность каналов в модели рассчитывалась в соответствии с реальными характеристиками. Ниже приведены данные о пропускной способности на каналах в период с 01.02.2008 по 25.02.2008. Графики с данными по сайтам — в приложении 1.

Сайт	Средняя пропускная способность (Мбит/с)
JINR	5,2
ИИЕР	4,9
ИТЕР	6,3
SINP	3,4

Модель тестировалась при изменении:

- 1) пропускной способности каналов;
- 2) размера файлов (увеличении размера файлов в два раза).

Все эти параметры прописывались в конфигурационных файлах.

РЕЗУЛЬТАТЫ

Как и следовало ожидать, при увеличении размеров файлов время выполнения задач увеличилось.

В таблицах представлены результаты работы модели — как изменяется время выполнения заданий в зависимости от изменения пропускной способности и размера файлов.

Моделировалась ситуация, когда на сайты копировалось следующее количество данных:

Сайт	Получено данных (Гбайт)
JINR	500
ИИЕР	340
ИТЕР	680
SINP	680

При изменении пропускной способности каналов и некотором варьировании фоновых нагрузок на канал время передачи заметно улучшалось.

Пропускная способность каналов, Мбит/с	4–6	8–10	12–14
Время выполнения заданий, ч	13,5	12,5	10,8

Увеличение количества передаваемых файлов в два раза дает следующие результаты:

Пропускная способность каналов, Мбит/с	4–6	8–10	12–14
Время выполнения заданий, ч	17,5	15,8	12,6

Реальные объемы передачи данных представлены ниже (более подробная информация в приложении 2):

Сайт	Средняя пропускная способность, Мбит/с	Среднее количество переданных данных, Гбайт/сут	Максимальное количество переданных, Тбайт/сут
JINR	5,2	400	1,4
INEP	4,9	450	0,8
ITEP	6,3	600	1,8
SINP	3,4	250	0,65

Видно, что при пропускной способности 6–12 Мбит/с на передачу 500–700 Гбайт данных потребуется 12–14 ч — такая ситуация близка к реальности. Это можно наблюдать на графике, приведенном в приложении 3. Однако количество переданных за час данных может различаться от нескольких гигабайт до сотен гигабайт. Предложенная ситуация отражает лишь один из вариантов развития событий. Можно предположить, что дальнейшее увеличение объемов передачи данных будет возможным при увеличении пропускной способности каналов и стабильности передач.

В дальнейшем необходимо развитие модели: добавление сайтов, тестирование при различных нагрузках и алгоритмах. Доступность ресурсов может изменяться в процессе работы, поэтому моделирование сбоев будет очень интересным для изучения.

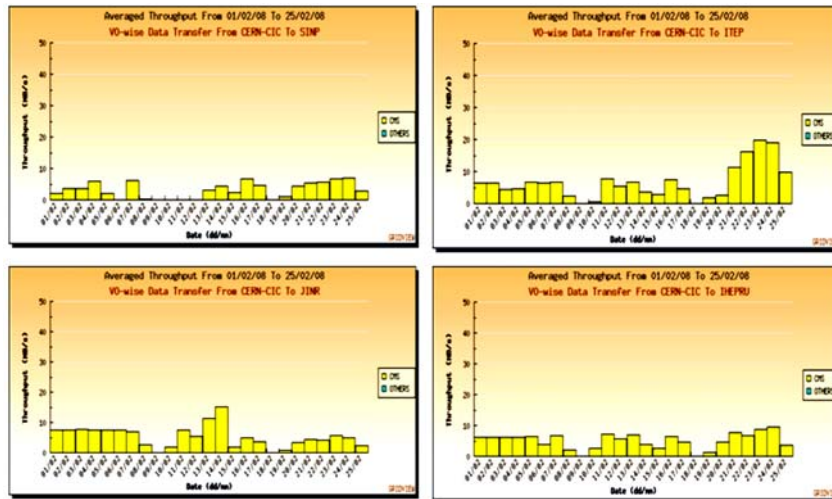
ЗАКЛЮЧЕНИЕ

В работе рассмотрены основные задачи, решаемые системами моделирования DataGrid. Для решения этих задач возможно применение пакетов моделирования OptorSim и GridSim. С помощью OptorSim создана модель реальной DataGrid структуры.

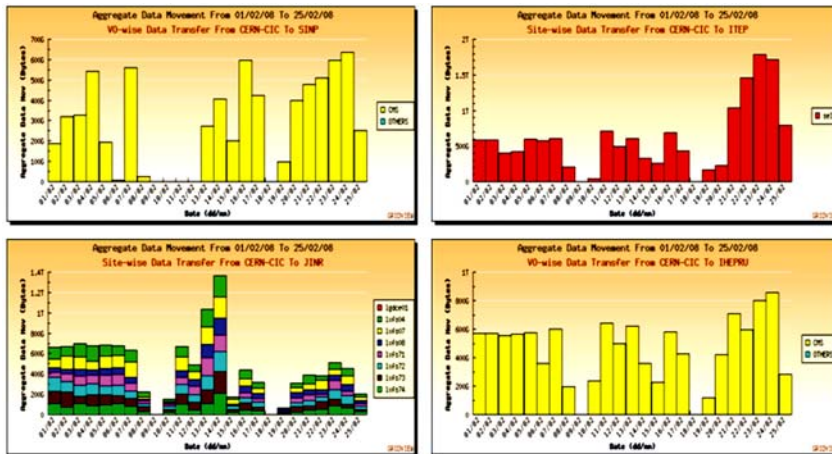
На примере моделирования участка CERN–RDIG показано применение OptorSim для моделирования DataGrid.

Авторы благодарны коллегам Jamie Shiers, Gavin Mccance, Paolo Tedesco и Steve Traulen за многочисленные и полезные обсуждения, а также за интересное и плодотворное сотрудничество — Владимиру Витальевичу Ужинскому.

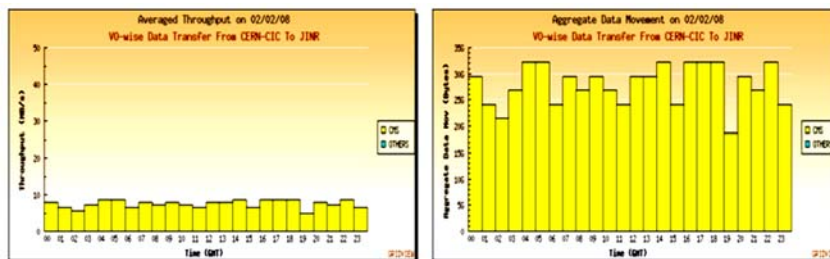
Приложение 1. Средняя пропускная способность каналов между ЦЕРН и сайтами RDIG (с 01.02.2008 по 25.02.2008)



Приложение 2. Среднее количество переданных данных между ЦЕРН и сайтами RDIG (с 01.02.2008 по 25.02.2008)



Приложение 3. Пропускная способность канала CERN-JINR и количество переданных данных за 02.02.2008



ЛИТЕРАТУРА

1. <http://lhc.web.cern.ch/lhc/>
2. www.cern.ch/lcg
3. *Sulistio A., Cibej U., Robic B., Buyya R.* A Toolkit for Modelling and Simulation of Data Grids with Integration of Data Storage, Replication and Analysis. http://gridbus.org/reports/datagrid_fgcs.pdf, 2006.
4. egee-rdig.ru
5. <http://www-csag.ucsd.edu/projects/grid/microgrid.html>
6. <http://gcl.ucsd.edu/simgrid>
7. <http://edg-wp2.web.cern.ch/edg-wp2/optimization/optorsim.html>
8. <http://www.gridbus.org/gridsim/>
9. *Cameron David G. et al.* OptorSim v2.1: Installation and User Guide, OptorSim official Sourceforge website:<http://sourceforge.net/projects/optorsim>, Oct 2006.
10. *Caminero A., Sulistio A., Caminero B., Carrion C., and Buyya R.* Extending Grid-Sim with an Architecture for Failure Detection. <http://www.gridbus.org/~raj/papers/gridsim-failure-detection-icpads2007.pdf>; <http://www.gridbus.org/~raj/papers/gridsim-failure-detection-icpads2007.pdf>, 2007.
11. *Buyya R., Abramson D., Giddy J. Nimrod-G.* An Architecture for a Resource Management and Scheduling system in a Global Computational Grid // Proc. of the 4th International Conference and Exhibition on High Performance Computing in Asia-Pacific Region, Beijing, China, 2000.
12. <http://www.geant2.net/server/show/nav.00d007009>
13. <http://www.egee-rdig.ru/>
14. <http://www.osp.ru/os/2008/02/4926522/>
15. <http://egee-jra1-dm.web.cern.ch/egee-jra1-dm/FTS/#transferjobs>

Получено 28 мая 2008 г.

Редактор *М. И. Зарубина*

Подписано в печать 21.10.2008.

Формат 60 × 90/16. Бумага офсетная. Печать офсетная.

Усл. печ. л. 0,75. Уч.-изд. л. 0,90. Тираж 310 экз. Заказ № 56388.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.

E-mail: publish@jinr.ru

www.jinr.ru/publish/